额尔齐斯河库威站日尺度的降雨融雪径流模拟

赵文龙[1,2]， 吕海深[1,2]， 朱永华[1,2]， 刘 涵[1,2]， 吴卓珺[1,2]
（1. 河海大学水灾害防御全国重点实验室，江苏 南京 210098；
2. 河海大学水文水资源学院，江苏 南京 210098）

摘 要：额尔齐斯河流域受地理条件的影响，流域内水文气象站点较少，基础资料匮乏，而融雪洪水在该流域的汛期及水资源管理上有着较大影响。本研究通过应用降水和气温的再分析产品及AVHRR积雪数据，利用K-means聚类法进行不同径流时期特点的划分，并在不同时期构建相应SRM+LSTM模型，并使用2009年数据及2023年实地观测的径流数据进行验证。结果表明：再分析产品CMFD能够较好地应用于额尔齐斯河流域，并能根据降水、温度、积雪及径流间的关系得到不同径流划分时期，即12月11日—次年4月10日为积雪退水期、4月11日—8月10日为融雪降水产流期、8月11日为降水产流期。SRM模型模拟效果较差，大部分径流纳什效率系数（NSE）<0；而SRM+LSTM模型能够较好地模拟该流域的不同时期的径流，决定系数R^2均能达到0.5以上，纳什效率系数也能达到0.5以上，证明SRM+LSTM模型能够较好地应用于该地区，精度较高。

关键词：K-means聚类法；SRM模型；LSTM模型；径流模拟；额尔齐斯河

融雪径流是我国西北阿勒泰地区重要的淡水资源。但在气候变暖、春季融雪较快的条件下，其季节性水文循环特征可能会发生变化[1]。春季积雪强烈消融期间，极易引发融雪型洪水，给当地的农牧业生产和人民的生活都带来严重影响和经济损失[2]。全球气候变化大背景下，额尔齐斯河流域融雪型洪水频次呈增大趋势，造成巨大洪灾损失[3]。因此，迫切需要构建额尔齐斯河流域融雪径流模拟及预报的模型，这不仅有助于预报流域融雪型洪水、理解其水文特征，也为未来的水资源管理提供重要参考。

我国西北干旱地区由于气象站与水文站点较少，积雪、洪水等资料匮乏，但也已经有了相应研究，陈庭兴等[4]在西营河九条岭水文站基于GEV分布进行洪水特征分析，很好地拟合了该站的最大日流量序列；郭晨煜等[5]基于LH-OAT方法研究WRF-Hydro模型参数在河西内陆河流域的参数敏感性，并结合典型洪水分析各参数对洪水过程线的影响；孙铭悦等[6]在呼图壁河研究了两套气象数据在资料匮乏地区的适用性评估，并应用HBV模型模拟融雪径流得到格点数据集优于站点数据集的结论；张梅洁等[7]利用耦合融雪模块的新安江模型，研究了干旱区西营河流域日径流模拟过程，分析了该模型在流域上的适用性。

额尔齐斯河流域径流主要来源于融雪和降水，冬季积雪是该流域主要特征之一。春季随天气转暖，冰雪融水和春季降水的共同补给下常形成春汛，夏秋季径流逐渐减少，冬季流量更小，径流季节性特征明显。因此，对于该流域可以考虑对应季节进行模型构建，根据产流来源划分产流时期，减小季节性影响，以此提高模型模拟的精度。目前，已有部分学者对融雪期及降雨期的划分进行过研究，李彬权等[8]在大渡河进行不同方法的径流时期划分，确定了最终融雪期划分方案；杨春辉等[9]在嫩江上游流域研究划分出融雪期及融冻期，并分别进行产流分析；田琳[10]采用ECK递归数字滤波法对第二松花江河源段河道径流基流部分进行分割用以进行融雪时段划分，并在所划分融雪时段进行了径流模拟。

目前，对融雪径流的预测已经有应用较广的模型，如SWAT模型[11]、VIC模型[12]、SRM模型[13]等。

收稿日期：2024-03-29； **修订日期**：2024-06-27
基金项目：国家重点研发计划项目（2019YFC1510504）
作者简介：赵文龙（2000-），男，硕士研究生，主要研究方向为融雪径流模拟. E-mail: 221301010052@hhu.edu.cn
通讯作者：吕海深. E-mail: lvhaishen@hhu.edu.cn

然而,基于物理的分布式模型SWAT模型与VIC模型结构复杂,所需数据信息较多,对于额尔齐斯河这种缺乏数据资料的流域,没有足够的数据。SRM(Snowmelt Runoff Model)模型由瑞典科学家Martinec[14]于1975年首次提出,是在一定程度上考虑了融雪过程中物理现象的概念性模型。由于模型需要的数据较少,在实测资料较少的地区得到了广泛应用。如Siemens等[15]将融雪径流模型(SRM)与MODIS相结合,远程感知积雪观测,模拟加拿大西部阿萨巴斯卡河流域上游融雪径流;谢顺平等[16]在融雪季节应用SRM模型模拟了中国西北部玛纳斯河流域的融雪径流;Sharma等[17]在喜马拉雅地区中部的尼泊尔喜马拉雅山应用SRM模型,利用卫星数据研究融雪对布迪甘达基河流域(BGRB)总流量的贡献,并得到了较好的结果。

此外,经过越来越多的学者不断探索,单一的水文模型目前已经较难满足现有流域的径流模拟等水文研究。随着计算机技术的发展,其计算效率逐步提高,并逐步应用于多个领域,在水文学领域机器学习成为了研究重点。在径流模拟相关研究中,长短时记忆(Long Short-Term Memory,LSTM)神经网络在长时间序列数据学习训练中展现出超强的数据学习能力[18-19]。水文数据的高复杂和非线性时间序列的特点,让长短时记忆神经网络在径流模拟等水文研究上有更好的优势。Garg等[20]采用长短期记忆(LSTM)模型,研究了印度戈达瓦里河流域的洪水径流并得到了较好的结果。殷兆凯等[21]应用长短时记忆神经网络模型(LSTM)对不同预见期的径流预报分别建立了流域降雨径流模型,得出了短期预报精度有所提高的结论。殷仕明等[22]提出了基于迁移学习的长短时记忆神经网络(LSTM)水文模型,并在嘉陵江、乌江和岷江流域得到了更高精度的模拟径流。但是,流域的物理规律和特征还是需要具有概念性特征的水文模型参与[23],才能够获得更贴近研究区域的径流模拟及预报模型。Chen等[24]在九龙河流域应用SWAT模型与LSTM模型进行耦合,并得出该模型在数据稀缺地区能够较好地模拟径流的结论;Xiao等[25]在四川省嘉陵河流域应用TOPMODEL模型与LSTM模型提高了模拟径流的可靠性和准确性。

本文根据高黎明等[26]的研究利用CMFD气象驱动数据以及AVHRR积雪数据[27],结合径流数据根据李彬权等[8]的研究,利用K-means聚类法进行不同时期径流划分,减小模型径流模拟季节性影响,并将驱动SRM模型进行额尔齐斯河上游流域的径流模拟,随后将得到的模拟径流与气象数据通过LSTM模型进行学习得到新的径流模拟,并应用2009年观测数据与2023年实测的径流观测数据验证模型精度。与以往研究不同,应用SRM模型通过划分不同径流时期,减少季节性影响,从而提高模型模拟效果;并用水文模型与机器学习相结合,以达到资料较少地区提高模拟径流精度的效果。

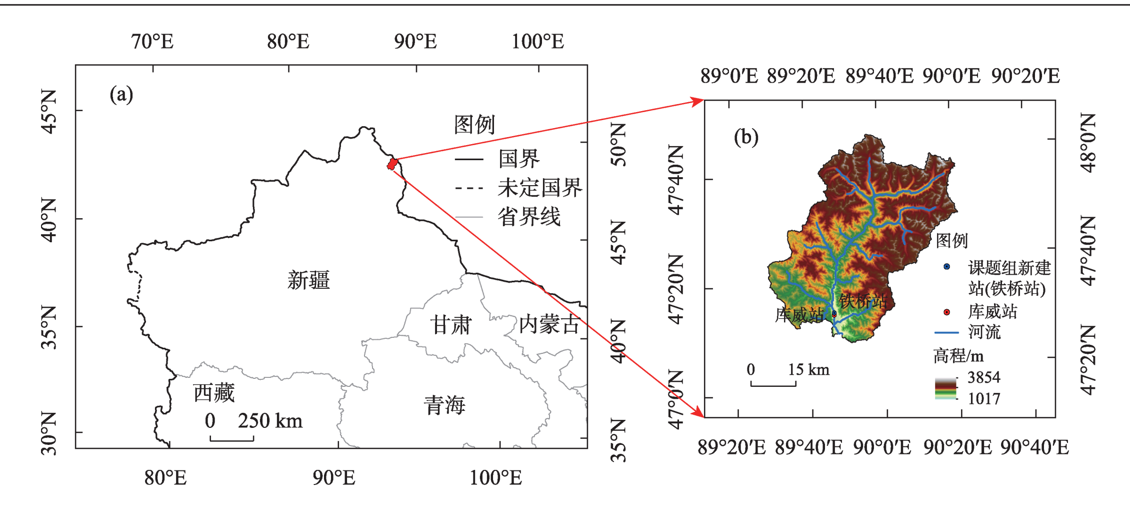
1 数据与方法

1.1 研究区概况

研究区域为卡依尔特斯河库威站及课题组新建站(铁桥站)以上的流域(图1),卡依尔特斯河是额尔齐斯河流域的一条支流,额尔齐斯河流域位于新疆阿勒泰地区北部,地理位置介于85°31′~90°29′E,46°50′~49°09′N之间。河流发源于阿尔泰山南麓,上游由库依尔特斯河和卡依尔特斯河汇集而成,下游在国界线处汇入俄罗斯境内的鄂毕河。额尔齐斯河流域地处温带大陆性干旱与半干旱气候区,上游径流主要靠融雪、融冰和降水补给;下游主要来源于融雪、降水和壤中水。上游汛期始于4月,大汛多在4—6月;下游汛期为5月末—10月,6月最大,约占全年的50%。研究区域流域面积为2494 km^2。

1.2 数据来源

本文所使用的气温、降水历史气象数据来自于中国区域地面气象要素驱动数据集(China Meteorological Forcing Dataset,CMFD),该数据集是以国际上现有的Princeton再分析资料、GLDAS资料、GEWEX-SRB辐射资料,以及TRMM降水资料为背景场,融合了中国气象局常规气象观测数据制作而成。数据为NETCDF格式,时间分辨率为3 h,水平空间分辨率为0.1°。该数据集可用于水文建模、地表建模、土地数据同化和其他陆地建模,精度介于气象局观测数据和卫星遥感数据之间,优于国际上已有再分析数据的精度[28]。积雪数据来自于国家冰川冻土沙漠科学数据中心的AVHRR中国积雪物候数据集,时间覆盖1980—2020年。雪深数据来自时空三级环境大数据平台的中国雪深长时间序列数据集[29]。径流数据取自《中华人民共和国水文年鉴》额尔齐斯河库威站2007—2009年实测径流数



注：底图采用自然资源部标准地图制作，审图号为GS(2023)2767号，对底图边界无修改。

图1　研究区地形示意图

Fig. 1　Topographic map of the study area

据。2023年5月9日—6月19日气温、降水数据及实测径流数据来自课题组新建站（铁桥站），该时间段积雪数据及雪深数据取自GFS（全球预报系统）数据。DEM数据取自地理空间数据云的ASTGTM2数据。

1.3　SRM模型

融雪径流模型（SRM）用于模拟或预测高山积雪融水为径流主要来源的流域，通过模型可以得到流域的每日径流量[14]。其计算原理是将每天的融雪和降水所产生的径流量求和，将退水系数计算到退水流量中，得到每日的径流过程，其核心计算公式如下：

$$Q_{n+1} = \left[C_{Sn} \times a_n\left(T_n + \Delta T_n\right)S_n + C_{Rn}P_n\right] \frac{A \times 10000}{86400}\left(1 - K_{n+1}\right) + Q_n K_{n+1} \quad (1)$$

式中：Q为平均日径流量（$m^3 \cdot s^{-1}$）；C_R为降雨径流系数；C_S为融雪径流系数；a为度日因子（$cm \cdot ℃^{-1} \cdot d^{-1}$），表示单位度日因子的融雪深度；$T$为度日因子数（$℃ \cdot d^{-1}$）；$\Delta T$为根据气温直减率在不同高程分带进行气温插值后度日因子数的调整值（$℃ \cdot d^{-1}$）；S为积雪的覆盖率；P为降水形成的径流深（cm）；A为流域分带的面积（km^2）；K为退水系数，反映了每日融水能直接补给到径流量中的比重，一般通过分析历史流量数据确定退水系数；n为时间序列值。

径流系数解释了损失，即流域流量与融雪+降雨之和的比值。SRM模型中径流系数C分为融雪径流系数C_S和降雨径流系数C_R。融雪初期，雪面蒸发较小，径流损耗较小；当融雪达到一定阶段时，由于土壤的暴露和植被生长会导致径流系数的降低。

度日因子a（$cm \cdot ℃^{-1} \cdot d^{-1}$）将度日因子数$T$（$℃ \cdot d^{-1}$）转换为每日融雪深度$M$（cm）：

$$M = a \times T \quad (2)$$

在缺乏详细数据的情况下，可以从经验关系中获得度日因子：

$$a = 1.1 \times \frac{\rho_s}{\rho_w} \quad (3)$$

式中：ρ_s为积雪密度（$kg \cdot m^{-3}$）；ρ_w为水的密度（$kg \cdot m^{-3}$）。

温度直减率反映了温度随高程变化的影响，高程越高，温度越低。如果研究区具有很多不同高度的温度站，直减率可以从历史实测数据中确定。温度直减率还可以通过周围相似流域移用。

临界温度决定了测量或预测的降水是雨还是雪。SRM模型判断降水是降雨还是降雪，积雪是否融化，都依赖临界温度的控制。临界温度以下，降水为降雪，积雪在累积；临界温度以上，降水为降雨，积雪融化为径流。

退水系数是SRM的一个重要参数，它反映了每日融水能直接补给到径流量中的比重。一般通过分析历史流量数据确定退水系数。退水系数的计算公式如下：

$$K_{n+1} = x \times Q_n^{-y} \quad (4)$$

对公式取对数得到：

$$\log Q_{n+1} = (1 - y)\log Q_n + \log x \quad (5)$$

式中：Q_n为第n天的日均流量（$m^3 \cdot s^{-1}$）；x、y为流域特征参数，具体由流域自身特点决定。

SRM模型的驱动数据主要包括气温、降水及积雪覆盖面积。本研究中气温和降水由上述提到的CMFD数据集提供，积雪覆盖面积通过AVHRR数据结合ArcGIS处理高程分带计算得到。通过ArcGIS处理高程分带如表1所示。

表1 高程面积分布

Tab. 1 Elevation area distribution

分带	高程范围/m	平均高程/m	面积占比累计/%	分带面积/km^2
1	1017~1517	1267	2.4	59.86
2	1517~2017	1767	24.0	538.71
3	2017~2517	2267	60.0	897.84
4	2517~3017	2767	93.6	837.98
5	3017~3854	3436	100.0	159.62

1.4 LSTM模型

长短时记忆神经网络（Long Short-Term Memory，LSTM），由输入层、一个或多个隐藏层、以及输出层组成。其隐藏层中的神经元不仅能从输入层中接收信息，还可以接收神经元从上一个时刻所感知的信息[30]。图2展示了一个典型结构：

具体过程如下：

在t时刻首先计算遗忘门f_t，f_t可以决定从之前的状态中舍弃多少信息。计算公式如下：

$$f_t = \sigma\left(U_f x_t + W_f h_{t-1} + b_f\right) \tag{6}$$

式中：U_f，W_f和b_f为遗忘门的可调参数矩阵或向量，在神经网络训练过程中将被优化。σ为Sigmoid激活函数。

下一步计算输入门i_t。i_t将决定从新获取的信息中选择多少用以更新状态。其计算公式如下：

$$i_t = \sigma(U_i x_t + W_i h_{t-1} + b_i) \tag{7}$$

式中：U_i，W_i和b_i为遗忘门的可调参数矩阵或向量，在神经网络训练过程中将被优化。

新获取信息$\tilde{c}_t$计算公式如下：

$$\tilde{c}_t = \tanh\left(U_{\tilde{c}} x_t + W_{\tilde{c}} h_{t-1} + b_{\tilde{c}}\right) \tag{8}$$

式中：$U_{\tilde{c}}$，$W_{\tilde{c}}$和$b_{\tilde{c}}$为$\tilde{c}_t$的可调参数矩阵或向量，训练过程中可被优化。Tanh函数为双曲正切激活函数。

随后，用以上的结果更新神经元状态，计算公式为：

$$c_t = f_t \times c_{t-1} + i_t \times \tilde{c}_t \tag{9}$$

然后计算输出门o_t，o_t可以决定在t时刻有多少信息生成隐藏层状态变量h_t，计算公式为：

$$o_t = \sigma(U_o x_t + W_o h_{t-1} + b_o) \tag{10}$$

$$h_t = o_t \times \tanh(c_t) \tag{11}$$

式中：U_o，W_o和b_o为输出门的可调参数矩阵或向量，在神经网络训练过程中将被优化。

最终h_t传入输出层再经过计算得到模型在t时刻的最终输出结果。

1.5 SRM与LSTM耦合过程

在额尔齐斯河单独应用SRM模型时，由于用于模型参数率定的驱动数据和径流数据的长度有限，模型模拟效果不佳。在探索水文模型输出数据的后处理方面时，发现LSTM模型对于水文序列数据之间关系的探索上表现较好，从而探究是否可以将水文模型的输出进行优化。由于长短期记忆神经网络的输入较为灵活，考虑将水文模型模拟结果作为长短期记忆神经网络方法的输入因子，同时考虑

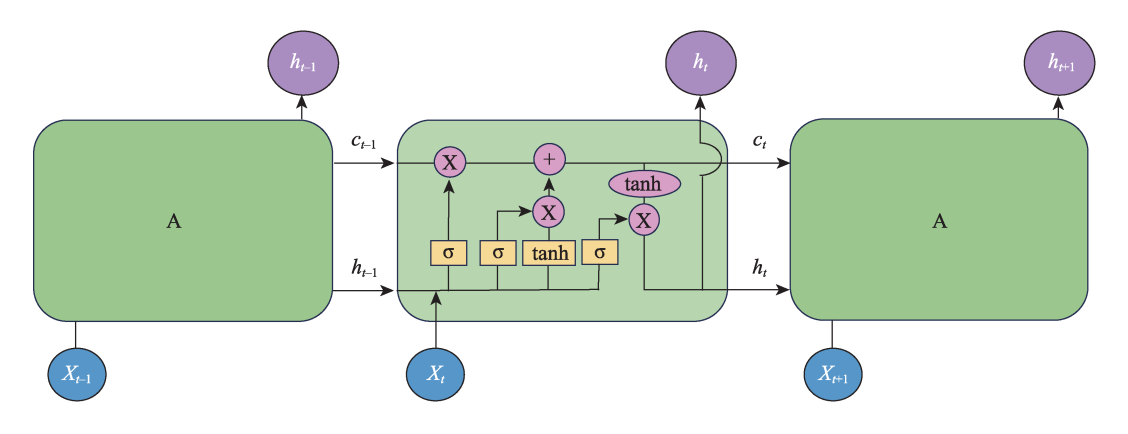


图2 LSTM模型及其内部结构

Fig. 2 LSTM model and its internal structure

降水和气温等其他气象数据可能对于径流产生的影响[23]。本研究模型的第一部分使用降水、气温和积雪数据输入于SRM模型，得到模拟径流；第二部分使用SRM模型模拟的径流结果，结合气温、降水及积雪覆盖数据和雪深数据，共同输入于LSTM模型，最后得出提高精度的模拟径流。

1.6 K-means聚类法

额尔齐斯河流域径流主要来源于融雪和降水，冬季积雪是该流域主要特征之一。春季随天气转暖，冰雪融水和春季降水的共同补给下常形成春汛，夏秋季径流逐渐减少，冬季流量更小，径流季节性特征明显。因此，对于该流域可以考虑对应季节进行模型构建，根据产流来源划分几个产流时期，减小季节性影响，以此提高模型模拟的精度。

聚类为一种无监督的分类方法，即将由多个独立样本构成的数据集经相似性指标判定进而划分为若干个类簇的过程。基于聚类思想的水文分期问题通过对连续时间内的特征指标进行相似性计算，从而得到携带有分类标签的各时段分期结果。K-means聚类方法作为主要的聚类方法，可以快速将其他个案归类到相应的类别，适合大样本数据的聚类。在对长时间序列水文数据集进行聚类分析时，算法聚类较高效且聚类效果较好[31]。

K-means聚类法的准则函数F计算公式为：

$$F = \sum_{x \in X} \min \| x - C_i \|^2 \tag{12}$$

式中：X为样本集合；x为类中样本位置矢量；C_i为类的中心点位置矢量。

本研究选取实测流量、降水、温度、积雪覆盖、雪深的多年旬均值作为该方法的输入项。定义K的取值范围为[2, 10]，最佳K值则通过计算轮廓系数S_{C_i}选取，该指标在−1~1之间，当轮廓系数达到最大值时，表示同类之间最为接近，此时聚类效果最优，具体计算公式为：

$$S_{C_i} = \frac{(b_i - a_i)}{\max(a_i, b_i)} \tag{13}$$

式中：i为样本点，i=1, 2, …, N；a_i为样本点i在同类别内到其他点的平均距离；b_i为样本点到最近不同类别中样本的平均距离。

1.7 模型精度评价指标

本研究所使用的评价标准为决定系数R^2和纳什效率系数NSE，计算公式分别为：

$$R^2 = \left[\frac{\sum_{i=1}^{n} (Q_i - \bar{Q}_i)(Q'_i - \overline{Q'_i})}{\sqrt{\sum_{i=1}^{n} (Q_i - \bar{Q}_i)^2} \sqrt{\sum_{i=1}^{n} (Q'_i - \overline{Q'_i})^2}} \right]^2 \tag{14}$$

$$NSE = 1 - \frac{\sum_{i=1}^{n} (Q_i - Q'_i)^2}{\sum_{i=1}^{n} (Q_i - \overline{Q'_i})^2} \tag{15}$$

式中：Q_i为实测日径流量；Q'_i为模拟日径流量；$\bar{Q}_i$为实测平均径流量；$\overline{Q'_i}$为模拟平均径流量；n为模拟期天数。决定系数R^2的取值范围在[0, 1]，越接近于1，说明模拟径流过程与实测径流过程越接近，模拟效果越好。NSE取值范围为[−∞, 1]，越接近1，说明模型效果好。

2 结果与分析

2.1 不同时期径流划分

为增强K-means聚类方法划分径流时期的可靠性，将本研究区域的日尺度CMFD气温数据、AVHRR积雪覆盖数据、中国雪深长时间序列数据集绘制时间序列图，得到2007—2009年温度、积雪覆盖率及雪深数据时间序列结果（图3）。

结合图3分析，总体来看，积雪覆盖率在每年11月开始至次年4月均为100%，此时，雪深虽然也在逐渐增长，但到次年2月就停止增长，此后开始减小；温度在11月至次年3月中旬在0 ℃以下，其余时间升至0 ℃以上。积雪覆盖率在4月上旬开始逐渐减小，并到6月降至为0；温度逐渐上升，雪深也在下降，并于6月降至为0。从6月开始至10月上旬，此时积雪覆盖率及雪深为0，温度在9月开始逐渐下降，到10月中下旬降至为0 ℃，而10月中旬至11月上旬，积雪覆盖率增长较快，此时温度在0 ℃以下，雪深也开始增长，但增速较慢。

根据K-means聚类法原理得到不同聚类数对应轮廓系数结果如图4所示。

轮廓系数越高，分类情况越好。由于当K为2时分类情况最好，K为3或5时次之，故本应选择K为2时所分结果。但结合图3所示的温度、积雪覆盖以及雪深的分析结果来看，结合实际降水径流情况，认为得到产流来源主要为融雪及降水；但由于冬季仅有降水且无融雪，导致径流更小，故认为K为2时在该流域代表性不强。而当K为5时，部分划分

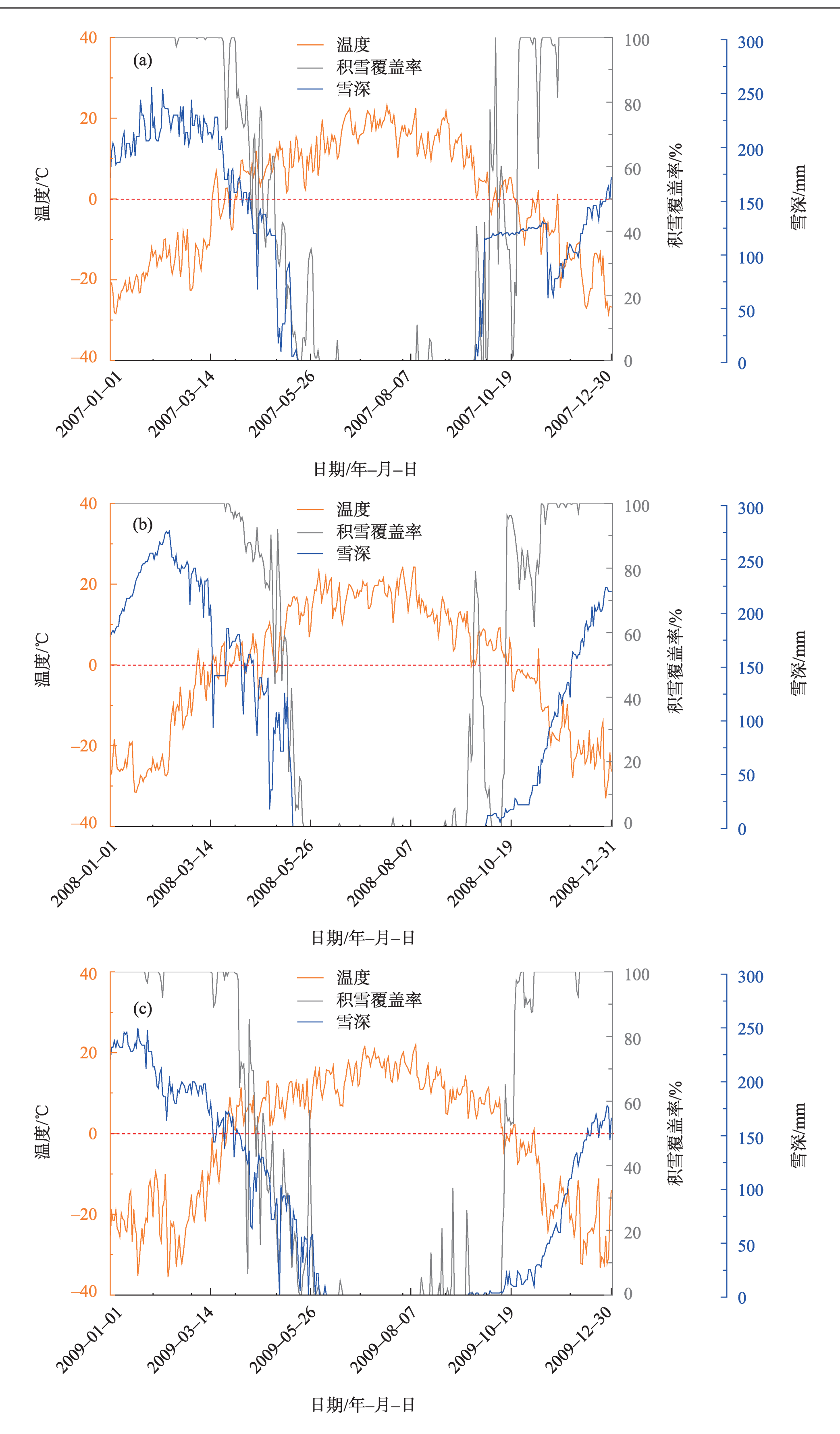


图3 2007—2009年日尺度温度、积雪覆盖率及积雪深度时间序列

Fig. 3 Time series of diurnal temperature, snow cover and snow depth from 2007 to 2009

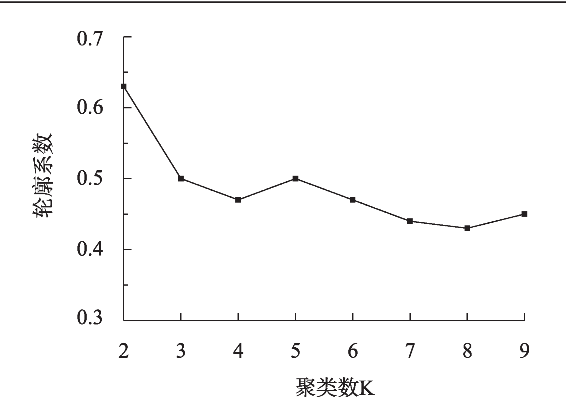


图4 K-means聚类法不同聚类数对应轮廓系数

Fig. 4 The corresponding contour coefficients of different clustering numbers in the K-means clustering method

时期时间过短,无法体现明显径流特征。当K为3时,可以凸显出产流来源及径流特征,且较为符合各时期实际径流特点,因此得到如表2所示的结果。

表2 K-means聚类法所分时期

Tab. 2 Periods of the K-means clustering method

时期	时间范围
积雪退水期	12月11日—次年4月10日
融雪降水产流期	4月11日—8月10日
降水产流期	8月11日—12月10日

结合图3分析来看,积雪退水期指的是当积雪覆盖率为接近100%时,且径流量基本处于低流量值时的径流时期;融雪降水产流期指的是积雪消融期间,主要由融雪和降水共同补给下产生的径流时期,该时期流量较大;降水产流期指的是积雪基本消融完成,基本仅由降水补给径流的径流时期,此时流量值主要影响因素为降水。由于积雪退水期的径流在比较时缺少部分资料,故在本研究中以当年1月1日—4月10日进行分析。

2.2 模型变量的确定

2.2.1 SRM模型参数的确定 在流域气象水文特征的基础上,结合金浩宇等[32]研究的参数基础上,通过调整获取径流系数值,结合本研究区具体情况,确定降雨径流系数和降雪径流系数(表3)。

对于度日因子,参考阿尔泰山融雪期不同下垫面积雪特性观测与分析研究成果,借鉴张伟等[33]在融雪期积雪密度研究成果,结合库威水文站控制区域实际情况,确定度日因子为0.1825~0.2552。温度直减率的选择采用李巧媛等[34]的研究,在海拔较高的干旱半干旱地区,温度直减率为0.65 ℃·(100m)$^{-1}$,这与国际通用的结果相同。临界温度的设定直接影响了产流的计算,本文的临界温度值取0 ℃。衰退系数选用库威站2007—2009年径流数据,并将$Q_{n+1}≥Q_n$时的数据剔除,绘制双对数散点图,拟合函数后根据公式得到x为1.09,y为0.05。

表3 径流系数

Tab. 3 Runoff coefficients

时期	降雨径流系数	降雪径流系数
积雪退水期	0.1	0.8
融雪降水产流期	0.8	0.8
降水产流期	0.5	0.1

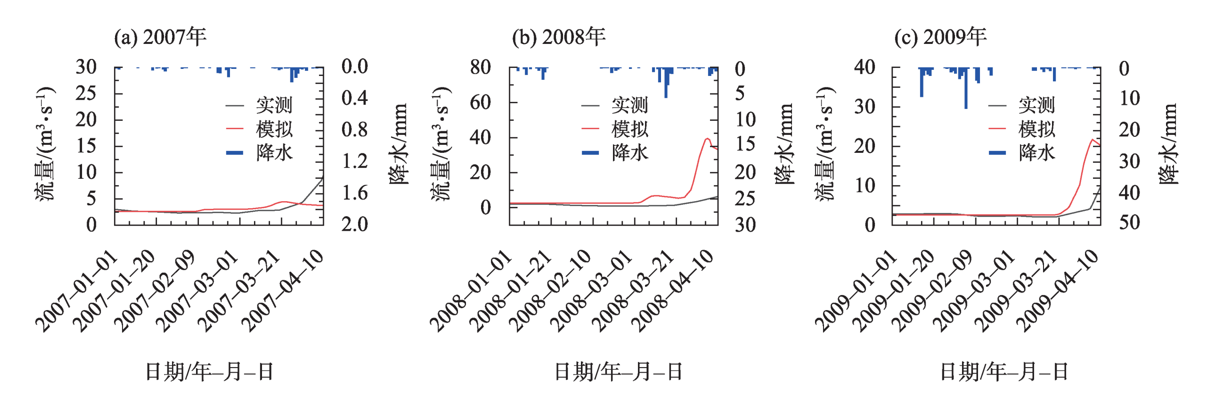
2.2.2 LSTM模型超参数的确定 本研究除了将SRM模型输出模拟流量及所用其他气象数据为输入外,在LSTM模型的建立过程中还有一些超参数需要提前确定。这些超参数将会决定LSTM模型的结构和训练等各个方面。其中隐藏层层数和隐藏层内神经元数量(n)将会决定LSTM模型的结构。本研究中通过测试将隐藏层层数确定为3层,隐藏层内神经元数量(n)作为LSTM模型复杂程度的唯一代表,将分别取值为2、4、16、128、256及512,最终选定数量n为256;在训练方面,将学习率设定为0.001,最大训练代数为500代。模型优化算法方面使用自适应矩估计(Adaptive Moment Estimation,ADAM)算法。本研究选择均方误差(MSE)作为LSTM模型优化过程中的目标函数。

2.3 SRM模型模拟结果分析

对额尔齐斯河流域的支流卡依尔特斯河,应用K-means聚类法获得不同径流划分时期。以2007—2008年实测径流数据作为参数率定年,2009年以及2023年5月9日—6月19日时间为验证年。得到SRM模型模拟结果如图5、图6、图7所示。

由图5可以看出,积雪退水期的径流量普遍偏低,但模拟情况较好,结合图3的温度、积雪覆盖及雪深时间序列分布来看,主要原因在于该时期降水较少,大部分积雪并未融化来补充径流导致。图5a中2007年的径流量由于降水资料值较小,融雪尚未开始,该时期末的径流量虽有上升但模拟效果并不好,图5b、图5c中2008年模拟期及2009年验证期的时期末模拟径流量均偏高,这可能由于降水和融雪相对2007年较多,从而导致的结果偏高。

从图6a可以看出,融雪降水产流期的模拟结果



注:(a)、(b)为率定期结果,(c)为验证期结果。下同。

图5 积雪退水期SRM模型径流模拟

Fig. 5 Runoff simulation of SRM model during snow retreat period

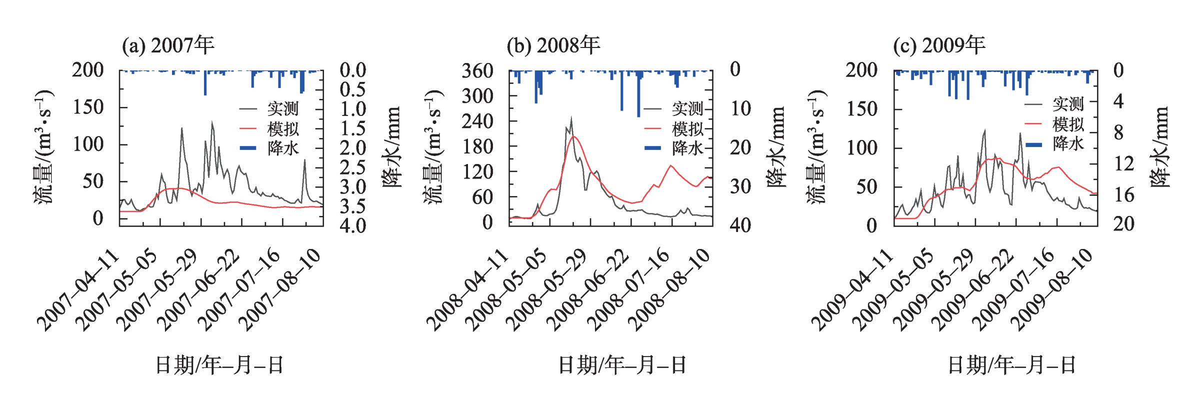


图6 融雪降水产流期SRM模型径流模拟

Fig. 6 Runoff simulation of SRM model during the runoff period of snowmelt precipitation

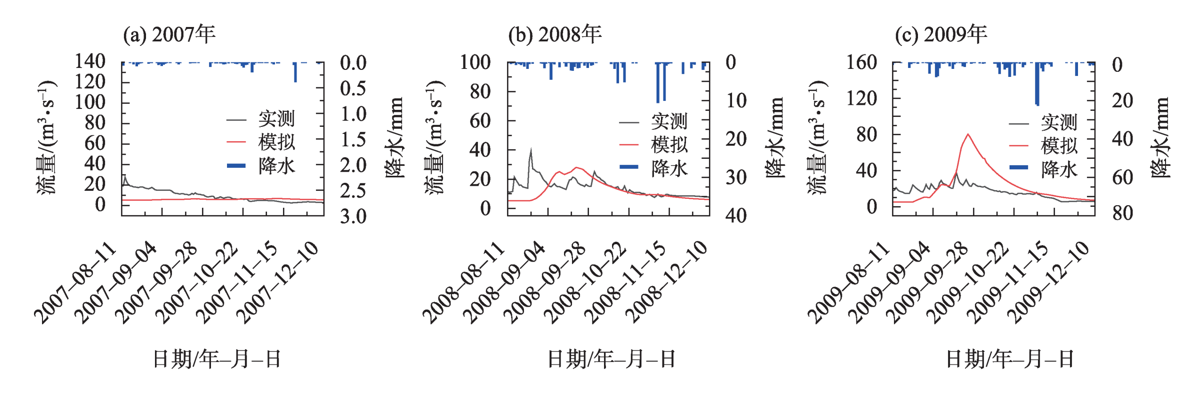


图7 降水产流期SRM模型径流模拟

Fig. 7 Runoff simulation of SRM model during precipitation runoff period

在2007年的径流模拟情况较差,分析认为可能由于降水资料偏低导致;而在图6b、图6c中2008年与2009年总体上洪峰模拟情况较好,但在该时期末所得到的模拟结果也偏高,分析可知2008年与2009年在该时期中后期存在大量降水,进而导致模拟结果偏高。

从图7分析在降水径流时期2007年由于冬季降水稀少,加之积雪不再融化,径流补给较少,整体上以退水期基流为主;2008年由于降水时间分布偏后,流量模拟情况也偏后,且降水量较上一年明显增加,使得洪峰流量有所上升;2009年降水情况与2008年类似,虽有洪峰但结果模拟偏高。

总体来说，SRM模型模拟效果精度较差，结合表4来看，很多洪水模拟纳什效率系数均<0，难以达到模拟较好的结果。分析发现，在模拟低流量时段时，模型模拟较为准确；并且在以往研究中，大部分SRM模型模拟时段为1 a及以上，这样在计算评价指标时，低流量时段的评价指标影响了总体的评价结果；而本研究流域中由于季节性特征明显，不同时期径流特征不同，模拟时段缩短，部分时段低流量较少，进而导致了计算评价指标时结果较差。本文也发现，在SRM模型中的融雪部分，对模型模拟结果影响较大。本研究中使用的以及目前的大多数积雪数据在该流域格网覆盖面大，由于高程分带较为精细，导致积雪数据存在较大误差影响，从而使模型结果与实际观测结果来看有较大误差，进而导致较多模拟结果纳什效率系数<0，难以得到较好的结果。

2.4 SRM+LSTM模型模拟结果分析

综上所述，SRM模型模拟结果虽然在退水期整体来看相对较好，但在降水产流期对于一些洪峰处有偏低或滞后的影响，在退水时期末也有偏高的影响，且纳什效率系数大部分都<0，精度较差。针对以上结果，结合数据与模型来看，SRM模型模拟结果精度较差，原因在于降水对于洪水的影响较小，而融雪占主要部分；并且数据量覆盖的时间范围较小；模型参数也多为经验参数，如温度直减率等，由于数据的匮乏也无法获得更加准确的数值。而通过将SRM模型模拟结果，结合气象数据进行LSTM模型模拟得到结果如图8、图9、图10所示。

经过LSTM模型的改进，从图8可以看出，在图8的积雪退水时期模拟结果更加接近实测径流，图8a的2007年模拟径流在时段末也有所上升，决定系数达到0.8以上，2008年与2009年的模拟径流也在时段末降至与实测值接近，总体纳什效率系数也符合模型模拟标准均>0.5，且2007年与2008年甚至达到了0.8以上。

从图9也可得出，在融雪降水径流时期，改进后的径流的趋势更加接近实际观测结果，除2007年洪峰值偏低外，2008年、2009年的洪峰流量改进结果

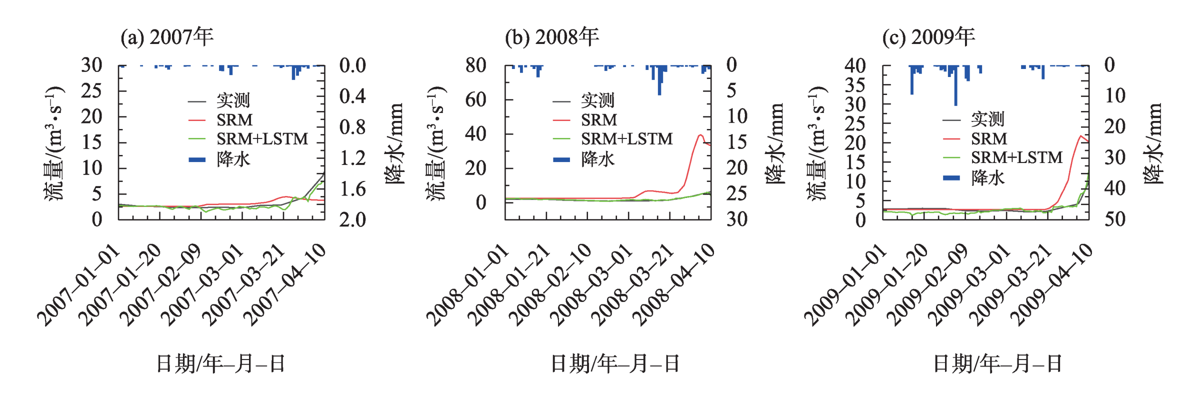


图8　积雪退水期SRM+LSTM径流模拟

Fig. 8　Simulation of SRM+LSTM runoff during the snow retreat period

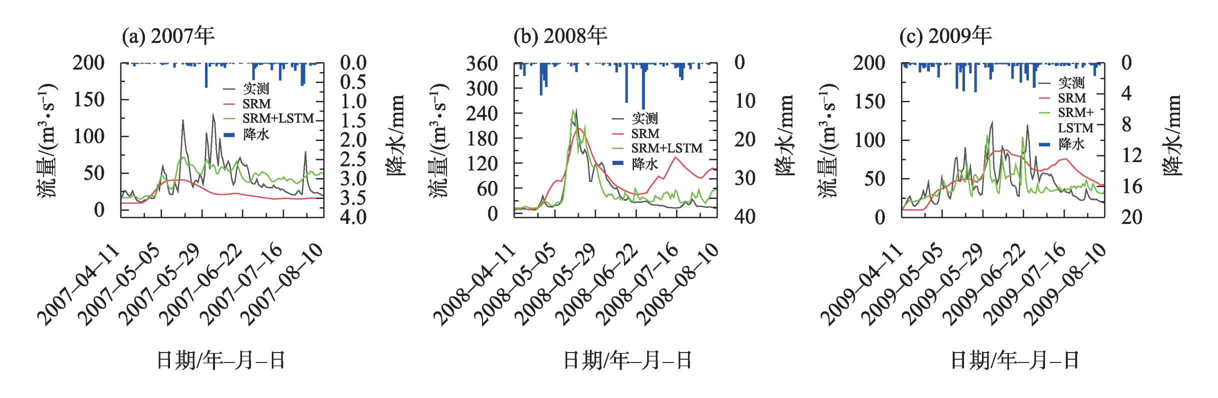


图9　融雪降水产流期SRM+LSTM径流模拟

Fig. 9　Simulation diagram of SRM+LSTM runoff during the runoff period of snowmelt precipitation

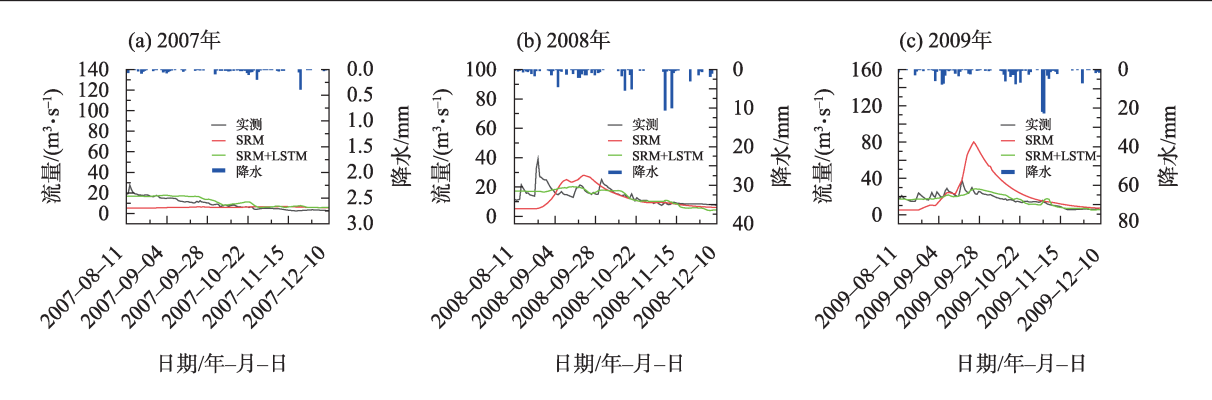


图10 降水产流期SRM+LSTM径流模拟

Fig. 10 Simulation diagram of SRM+LSTM runoff during precipitation runoff

较之前更好，时期末偏高的流量也被校正良好；2007年的确定性系数为0.441，2008年与2009年确定性系数也达到了0.3以上；2007年的结果可能由于之前SRM模型模拟结果较差以及输入数据资料的问题，导致最后改进的径流值不如2008年及2009年的结果；2007年与2009年的NSE达到0.2以上左右，而2008年的NSE达到0.8以上。

从图10结果来看，2007年的径流量由于降水偏少，仍然偏低，但相对SRM模型结果来说已有所改善，决定系数达到0.815，纳什效率系数也达到了0.646；2008年降水径流时期流量也更加接近实测值，但洪峰模拟结果仅相对SRM模型模拟结果更好，纳什效率系数为0.468；2009年的洪峰改进结果较好，与之前所得结果相比改进较大，NSE值也达到了0.5以上。

各时期的评价指标结果如表4所示。

根据图11以及图12综合分析来看，2023年6月初温度陡升，积雪覆盖及雪深降至0，证明雪水已融化，进而补给了径流量。结合图11来看，流量增大，SRM模型模拟结果虽有上升，但并未模拟出真正的洪峰流量，决定系数虽较高，但NSE仅为0.060，模型模拟效果较差，经过LSTM模型改进后，NSE提升至0.628，洪峰流量已有显著提高，模型模拟效果较好。

总的来说，通过引入LSTM模型，模型模拟精度有明显提高，原因认为是SRM模型所需气象要素类型较少，作为重要影响因素的融雪部分并未得到较好的展示，而增加雪深作为影响因素输入于LSTM模型中，提高了模型的模拟精度；而LSTM模型对于流量数据及气象数据的关联性有较强的学习能力，能够更好地模拟融雪径流，但仅靠该模型缺少物理因素的分析，而结合了SRM模型后，弥补了这个缺点，因此，耦合模型的模拟精度能够有所提高。上述结论也证明所建立的模型能够较好的模拟验证2009年与2023年的融雪径流。这也可以得出，本研究所得到的耦合模型在不依赖具有在额尔齐斯河高度适用性的CMFD数据时也能得到较好的结果。

表4 模拟结果评价指标

Tab. 4 Evaluation indicators of simulation results

时期		年份	SRM		SRM+LSTM	
			R^2	NSE	R^2	NSE
积雪退水期	率定期	2007年	0.278	0.266	0.922	0.878
		2008年	0.815	<0	0.925	0.902
融雪降水产流期	验证期	2009年	0.689	<0	0.817	0.555
	率定期	2007年	0.160	<0	0.441	0.422
		2008年	0.496	<0	0.834	0.814
降水产流期	验证期	2009年	0.334	0.082	0.314	0.277
	率定期	2007年	0.073	<0	0.815	0.646
		2008年	0.186	<0	0.513	0.468
融雪降水产流期	验证期	2009年	0.420	<0	0.753	0.749
	验证期	2023年	0.494	0.060	0.834	0.628

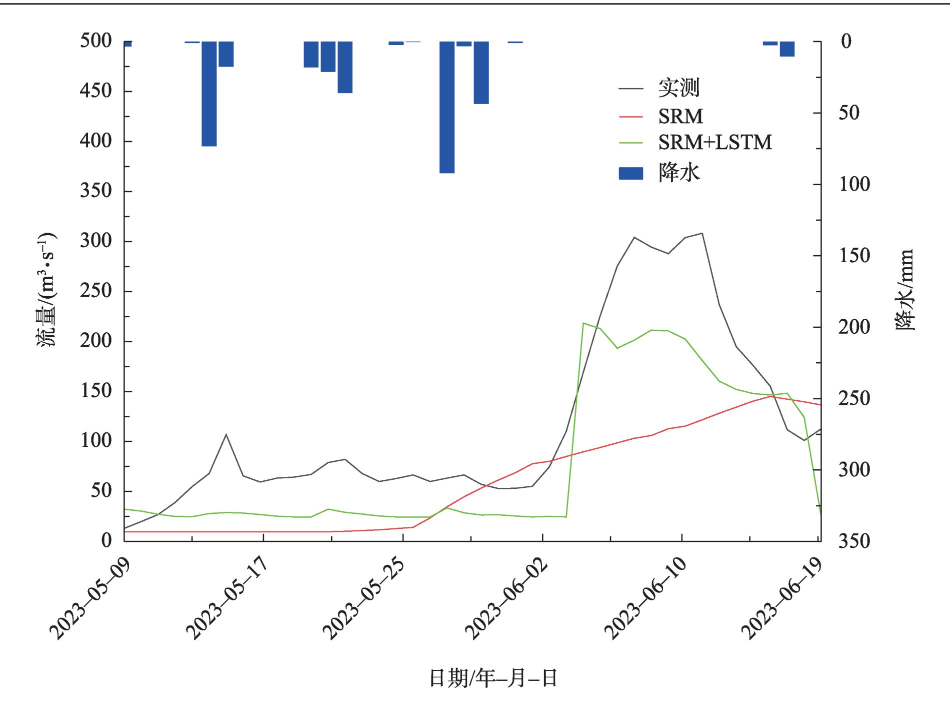


图11　2023年融雪降水产流期SRM+LSTM径流模拟

Fig. 11　Simulation of SRM+LSTM runoff during the runoff period of snowmelt precipitation in 2023

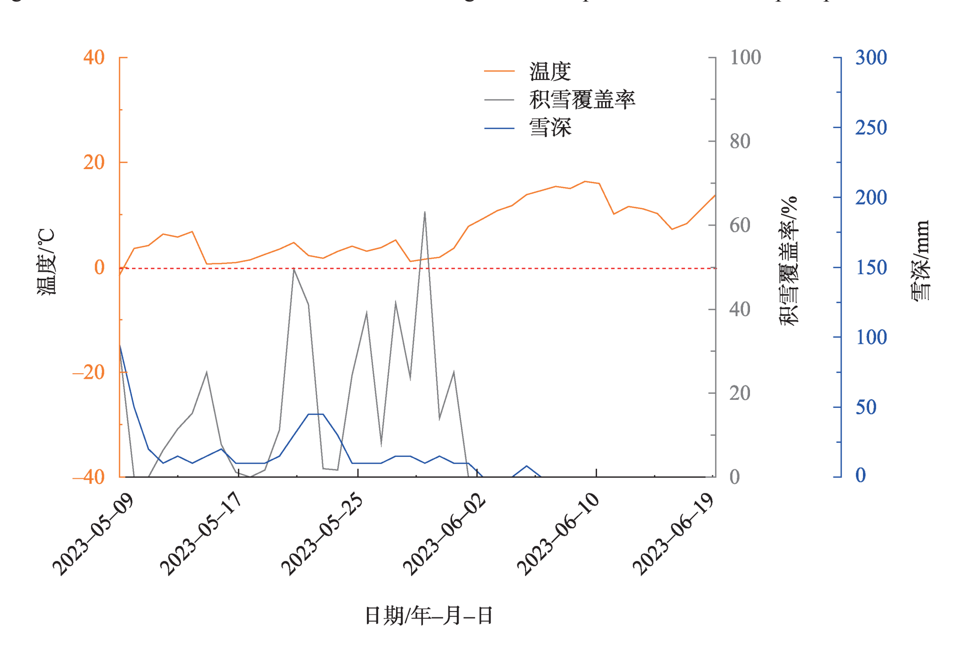


图12　2023年气温、积雪覆盖率、积雪深度时间序列

Fig. 12　Time series of temperature, snow cover, and snow depth in 2023

由于模型所需数据较少,便于建立等特点,在我国西北这种缺乏资料的干旱地区也能较好地模拟融雪径流。

3　结 论

本文通过应用CMFD数据及课题组新建站(铁桥站)所得2023年部分时期气象数据在额尔齐斯河流域构建SRM模型以及SRM+LSTM模型进行径流模拟,并用K-means聚类法进行不同径流时期的分类来进行短时期径流模拟,减少季节性影响,提高模拟精度,以决定系数与纳什效率系数为指标,得出以下结论:

(1) 额尔齐斯河流域受地理条件及经济社会的影响,流域内水文气象站点较少,基础资料匮乏,利

用再分析产品作为模型输入变量成为SRM+LSTM模型在额尔齐斯河流域应用的优势。CMFD数据能够有效应用在额尔齐斯河流域。

（2）K-means聚类法能够较好地反映温度、降水、径流、积雪之间的关系，并可以根据径流特点将其时间分为积雪退水期、融雪降水产流期及降水产流期。这对提高不同时期径流模拟精度有重要作用。

（3）SRM模型模拟结果较差，但SRM+LSTM模型整体模拟结果较好，2023年应用课题组新建站（铁桥站）的降水、气温数据用于所建模型中也能得到可靠结果。决定系数大部分都在0.5以上，NSE也达到了0.5以上。证明SRM+LSTM模型能够较好地应用于该流域。

参考文献（References）：

[1] Wu Xuejiao, Zhang Wei, Li Hongyi, et al. Analysis of seasonal snowmelt contribution using a distributed energy balance model for a river basin in the Altai Mountains of northwestern China[J]. Hydrological Processes, 2021, 35(3): e14046.

[2] 杨金明, 李诚志, 房世峰, 等. 新疆地区季节性融雪洪水模拟与预报研究[J]. 新疆大学学报(自然科学版), 2019, 36(1): 80–88. [Yang Jinming, Li Chengzhi, Fang Shifeng, et al. Simulation and forecasting of seasonal snowmelt flood in Xinjiang[J]. Journal of Xinjiang University (Natural Science Edition), 2019, 36(1): 80–88.]

[3] 余其鹰, 胡彩虹, 白云岗, 等. 新疆洪水预报预警中融雪径流模型应用进展[J]. 干旱区地理, 2023, 46(12): 1951–1962. [Yu Qiying, Hu Caihong, Bai Yungang, et al. Progress in the application of snowmelt runoff model in flood forecasting and early warning in Xinjiang[J]. Arid Land Geography, 2023, 46(12): 1951–1962.]

[4] 陈庭兴, 吕海深, 朱永华. 基于GEV分布的西营河流域洪水特性分析[J]. 干旱区研究, 2021, 38(6): 1563–1569. [Chen Tingxing, Lyu Haishen, Zhu Yonghua. Analysis of flood characteristics in Xiying River Basin based on GEV distribution[J]. Arid Zone Research, 2021, 38(6): 1563–1569.]

[5] 郭晨煜, 吕海深, 朱永华, 等. WRF-Hydro模型参数在河西内陆河流域的敏感性分析[J]. 水利水电科技进展, 2023, 43(6): 120–127. [Guo Chenyu, Lyu Haishen, Zhu Yonghua, et al. Sensitivity analysis of WRF-Hydro model parameters in hexi inland river basin[J]. Progress in Water Resources and Hydropower Science and Technology, 2023, 43(6): 120–127.]

[6] 孙铭悦, 吕海深, 朱永华, 等. 2套气象数据在资料缺乏地区的适用性评估——以呼图壁河流域为例[J]. 干旱区研究, 2022, 39(1): 94–103. [Sun Mingyue, Lyu Haishen, Zhu Yonghua, et al. Evaluation of the applicability of two sets of meteorological data in areas lacking data: A case study of Hutubi River Basin[J]. Arid Zone Research, 2022, 39(1): 94–103.]

[7] 张梅洁, 吕海深, 刘娣, 等. 耦合融雪的新安江模型在干旱区径流模拟研究[J]. 干旱区研究, 2022, 39(2): 379–387. [Zhang Meijie, Lyu Haishen, Liu Di, et al. Runoff simulation of Xin'anjiang model coupled with snow melting in arid region[J]. Arid Zone Research, 2022, 39(2): 379–387.]

[8] 李彬权, 荣伟光, 吴亚琪, 等. 大渡河流域融雪期划分方法研究[J]. 水电能源科学, 2023, 41(5): 38–42. [Li Binquan, Rong Weiguang, Wu Yaqi, et al. Study on the division method of snowmelt period in Dadu River Basin[J]. Hydropower Energy Science, 2023, 41(5): 38–42.]

[9] 杨春辉, 张玥, 李峰平. 寒区春汛产流预报方法研究[J]. 中国农村水利水电, 2019(5): 43–46. [Yang Chunhui, Zhang Yue, Li Fengping. Research on spring flood runoff forecasting method in cold regions[J]. China Rural Water Resources and Hydropower, 2019(5): 43–46.]

[10] 田琳. 中温带地区春季融雪径流研究[D]. 长春: 吉林大学, 2019. [Tian Lin. Study on Spring Snowmelt Runoff in the Middle Temperate Zone[D]. Changchun: Jilin University, 2019.]

[11] 邢贞相, 金超群, 纪毅, 等. 基于SWAT模型与Copula修正的融雪径流模拟[J]. 东北农业大学学报, 2020, 51(6): 79–87. [Xing Zhenxiang, Jin Chaoqun, Ji Yi, et al. Simulation of snowmelt runoff based on SWAT model and Copula modification[J]. Journal of Northeast Agricultural University, 2020, 51(6): 79–87.]

[12] 汪伟, 方秀琴, 杜晓彤, 等. 基于VIC模型的柳江流域分布式水文模拟与应用[J]. 水土保持研究, 2020, 27(3): 328–335. [Wang Wei, Fang Xiuqin, Du Xiaotong, et al. Distributed hydrological simulation and application of Liujiang River Basin based on VIC model[J]. Soil and Water Conservation Research, 2020, 27(3): 328–335.]

[13] 陈心池, 张利平, 闪丽洁, 等. 新疆山区中小河流洪水预报模型及其应用[J]. 干旱区研究, 2017, 34(6): 1426–1435. [Chen Xinchi, Zhang Liping, Shan Lijie, et al. Flood forecasting model for small and medium-sized rivers in mountainous areas of Xinjiang and its application[J]. Arid Zone Research, 2017, 34(6): 1426–1435.]

[14] Martinec J. Snowmelt-runoff model for stream flow forecasts[J]. Hydrology Research, 1975, 6(3): 145–154.

[15] Siemens K, Dibike Y, Shrestha R R, et al. Runoff projection from an alpine watershed in western Canada: Application of a snowmelt runoff model[J]. Water, 2021, 13(9): 1199.

[16] 谢顺平, 都金康, 冯学智, 等. 融雪径流模型参数渐进式优化率定方法[J]. 南京大学学报(自然科学), 2015, 51(5): 1005–1013. [Xie Shunping, Du Jinkang, Feng Xuezhi, et al. Parameter progressive optimization calibration method for snowmelt runoff model[J]. Journal of Nanjing University (Natural Science), 2015, 51(5): 1005–1013.]

[17] Sharma T P P, Zhang Jiahua, Khanal N R, et al. Assimilation of

snowmelt runoff model (SRM) using satellite remote sensing data in Budhi Gandaki River Basin, Nepal[J]. Remote Sensing, 2020, 12(12): 1951.

[18] 刘攀, 郑雅莲, 谢康, 等. 水文水资源领域深度学习研究进展综述[J]. 人民长江, 2021, 52(10): 76-83. [Liu Pan, Zheng Yalian, Xie Kang, et al. Review of research progress on deep learning in the field of hydrology and water resources[J]. People's Yangtze River, 2021, 52(10): 76-83.]

[19] Xiang Zhongrun, Yan Jun, Demir I. A rainfall-runoff model with LSTM-based sequence-to-sequence learning[J]. Water resources research, 2020, 56(1): e2019WR025326.

[20] Garg N, Negi S, Nagar R, et al. Multivariate multi-step LSTM model for flood runoff prediction: A case study on the Godavari River Basin in India[J]. Journal of Water and Climate Change, 2023, 14(10): 3635-3647.

[21] 殷兆凯, 廖卫红, 王若佳, 等. 基于长短时记忆神经网络(LSTM)的降雨径流模拟及预报[J]. 南水北调与水利科技, 2019, 17(6): 1-9. [Yin Zhaokai, Liao Weihong, Wang Ruojia, et al. Rainfall runoff simulation and forecast based on long short-term memory neural network (LSTM)[J]. South-to-North Water Diversion and Water Science and Technology, 2019, 17(6): 1-9.]

[22] 殷仕明, 徐炜, 熊一橙, 等. 基于迁移学习的长短时记忆神经网络水文模型[J]. 水力发电学报, 2022, 41(6): 53-64. [Yin Shiming, Xu Wei, Xiong Yicheng, et al. Hydrological model of long short-term memory neural network based on transfer learning[J]. Journal of Hydroelectric Engineering, 2022, 41(6): 53-64.]

[23] Cui Zhen, Zhou Yanlai, Guo Shenglian, et al. A novel hybrid XAJ-LSTM model for multi-step-ahead flood forecasting[J]. Hydrology Research, 2021, 52(6): 1436-1454.

[24] Chen Shengyue, Huang Jinliang, Huang Jr-Chuang. Improving daily streamflow simulations for data-scarce watersheds using the coupled SWAT-LSTM approach[J]. Journal of Hydrology, 2023, 622: 129734.

[25] Xiao Qintai, Zhou Li, Xiang Xin, et al. Integration of hydrological model and time series model for improving the runoff simulation: A case study on BTOP model in Zhou River Basin, China[J]. Applied Sciences, 2022, 12(14): 6883.

[26] 高黎明, 张乐乐, 沈永平, 等. ERA-Interim和CMFD气象驱动数据在新疆额尔齐斯河流域的适用性评价[J]. 冰川冻土, 2022, 44(1): 179-187. [Gao Liming, Zhang Lele, Shen Yongping, et al. Applicability of ERA-Interim and CMFD weather-driven data in the Irtysh River Basin of Xinjiang[J]. Glaciology and Geocryology, 2022, 44(1): 179-187.]

[27] 郝晓华, 赵琴, 纪文政, 等. 1980—2020年AVHRR中国积雪物候数据集[J]. 中国科学数据(中英文网络版), 2022, 7(3): 50-59. [Hao Xiaohua, Zhao Qin, Ji Wenzheng, et al. AVHRR snow phenology dataset in China from 1980 to 2020[J]. Chinese Scientific Data Online, 2022, 7(3): 50-59.]

[28] He Jie, Yang Kun, Tang Wenjun, et al. The first high-resolution meteorological forcing dataset for land process studies over China [J]. Scientific Data, 2020, 7(1): 25.

[29] Chen Hongju, Yang Jianping, Ding Yongjian, et al. Simulation of daily snow depth data in China based on the NEX-GDDP[J]. Water, 2021, 13(24): 3599.

[30] Hochreiter Sepp, Schmidhuber Jurgen. Long short-term memory[J]. Neural computation, 1997, 9(8): 1735-1780.

[31] Jain A K, Murty M N, Flynn P J. Data clustering: A review[J]. ACM Computing Surveys (CSUR), 1999, 31(3): 264-323.

[32] 金浩宇, 鞠琴. SRM模型在尼洋河流域的应用研究[J]. 水文, 2019, 39(5): 19-24. [Jin Haoyu, Ju Qin. Application of SRM model in Niyang River Basin[J]. Journal of China Hydrology, 2019, 39(5): 19-24.]

[33] 张伟, 沈永平, 贺建桥, 等. 额尔齐斯河源区森林对春季融雪过程的影响评估[J]. 冰川冻土, 2014, 36(5): 1260-1270. [Zhang Wei, Shen Yongping, He Jianqiao, et al. Impact assessment of forests on spring snowmelt process in the headwaters of the Irtysh River[J]. Glaciology and Geocryology, 2014, 36(5): 1260-1270.]

[34] 李巧媛, 谢自楚. 高原区气温垂直递减率的分布及其特点分析——以青藏高原及其周边地区为例[J]. 石河子大学学报(自然科学版), 2006, 24(6): 719-723. [Li Qiaoyuan, Xie Zichu. Distribution and characteristics of vertical temperature decline rate in plateau region: A case study of the Tibetan Plateau and its surrounding areas[J]. Journal of Shihezi University (Natural Science Edition), 2006, 24(6): 719-723.]

Simulation of rainfall and snowmelt runoff on the daily scale of the Kuwei Station in the Irtysh River

ZHAO Wenlong[1,2], LYU Haishen[1,2], ZHU Yonghua[1,2], LIU Han[1,2], WU Zhuojun[1,2]

(1. National Key Laboratory of Water Disaster Prevention, Hohai University, Nanjing 210098, Jiangsu, China;

2. College of Hydrology and Water Resources, Hohai University, Nanjing 210098, Jiangsu, China)

Abstract: Due to geographical conditions, there are limited hydrometeorological stations and a lack of basic data in the Irtysh River Basin, and the snowmelt flood exerts a considerable effect on the flood season and water resources management in the basin. In this study, precipitation and temperature reanalysis products and AVHRR snow cover data were applied, the K-means clustering method was used to divide the characteristics of different runoff periods, the corresponding SRM+LSTM model in different periods was constructed, and the runoff data observed in the field in 2023 were used. Results showed that the reanalysis product CMFD can be well applied to the Irtysh River Basin according to precipitation and temperature. The relationship between snow cover and runoff was divided into different runoff periods, as follows: December 11th to April 10th of the following year was the snow retreat period, April 11th to August 10th was the snowmelt precipitation runoff period, and August 11th was the precipitation runoff period. The simulation effect of the SRM model was poor, and the Nash efficiency coefficient of most runoff was<0. The SRM+LSTM model could better simulate the runoff in different periods of the basin, the deterministic coefficient could reach>0.5, and the Nash efficiency coefficient NSE could also reach> 0.5, which confirms that the SRM+LSTM model can be better applied to the area with high accuracy.

Keywords: K-means clustering method; SRM model; LSTM model; runoff simulation; Irtysh River